DOES A RUBBER BASELINE GUARANTEE COST OVERRUNS ON DEFENSE ACQUISITION CONTRACTS?

David S. Christensen and James A. Gordon

Published in *Project Management Journal 29*:43-51 (September 1998)

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the Department of Defense of the U.S. Government.

BIOGRAPHY

David S. Christensen is an Associate Professor of Accounting at Southern Utah University. He received a Ph.D. from the University of Nebraska at Lincoln in 1987. With research and consulting interests in cost management, he has published extensively in numerous academic and professional journals, including the *Project Management Journal*, the *Journal of Cost Analysis*, the *Contract Management Journal*, and the *Acquisition Review Quarterly.* He is active in several professional associations, and is a Certified Management Accountant (CMA), a Certified Cost Analyst/Estimator (CCA/E), and a Certified Government Financial Manager (CGFM). Presently, he serves as an Associate Editor to the *Journal of Cost Analysis*.

James Gordon is an officer in the United States Air Force. He received a master's degree in systems management from the Air Force Institute of Technology in 1996. Presently, he is serving as the Chief of Logistics on the Evolved Expendable Launch Vehicle Program at Los Angeles Air Force Base, California.

DOES A RUBBER BASELINE GUARANTEE COST OVERRUNS ON DEFENSE ACQUISITION CONTRACTS?

ABSTRACT

A common assertion in the defense literature is that an unstable budget baseline contributes to cost overruns on defense acquisition contracts. Using cost performance data from over 400 defense acquisition contracts, we tested this assertion. The stability of the baseline was characterized by the number of significant changes to the budget, and by a statistical measure of the baseline's variability, the coefficient of variation. Cost performance was characterized by cost and schedule performance indices. Using two statistical methods we found no significant relationships between baseline instability and cost overruns. Further, these results were insensitive to the managing service, the buying activity, and the contract type. Changes on a defense contract are not compelling rationale for cost overruns. Other possible causal factors should be more closely examined.

INTRODUCTION

This study tests the popular assumption that changes to defense acquisition contracts cause cost overruns. Defense programs often take seven to ten years to develop and produce (Fox 1988:28). During this time, requirements will likely change. Often, a contract is not fully defined when it is awarded, and modifications become a way of life almost from the start. These changes are necessarily reflected in a contract's time-phased budget, which is termed a "rubber baseline" when the changes are large and frequent.

Observers often assert that a rubber baseline is a major cause of cost overruns (e.g., Gilbraeth 1986:37, Augustine *et al.* 1990:2). If this assertion can be supported, then policies that restrict modifications may produce fewer cost overruns. If the assertion cannot be supported, then frequent contract changes cannot be considered a credible explanation for cost overruns, and other possible causes of cost overruns should be more closely examined.

Cost overruns versus cost growth. A cost overrun is not the same as cost growth. A cost overrun is the difference between the budget for the completed work and the actual cost of the completed work. Cost growth is the difference between the *initial* budget for a project and its final cost. As additional work is added to a project, both the total budget and the total cost of that project necessarily increase. If the budget for the new work is adequate, there would not be a cost overrun, but there would be cost growth.

Figure 1 illustrates the difference between cost growth and a cost overrun for a completed, multi-year project. The initial budget for the project was $100 million, but because of several modifications, the final budget of the project increased to $112 million. The final cost of the completed project is $120 million. Using these data, the cost of the project grew by $20 million, while the final cost overrun was $8 million.

Distinguishing between cost growth and cost overruns can clarify the analysis of casual factors that contribute to each. While it is clear that increases in scope will necessarily change the budgetary baseline and increase the final cost of a project, it is not clear if those changes will also lead to cost overruns.

This study tests the assertion that changes in the budgetary baseline (primarily increases) lead to cost overruns. Based on an analysis of contract data from a database of completed defense contracts, the assertion could not be supported. Cost overruns are not correlated with baseline changes. This paper reviews the relevant literature, and describes the methodology and the results. The implications of these results conclude the paper.

LITERATURE REVIEW

There have been a number of studies (e.g., Drezner *et al.* 1993, Tyson *et al.* 1994, Czelusniak and Rodgers, 1997) which suggest that baseline volatility is a persistent characteristic on many defense programs, and accounts for a large portion of cost growth. After adjusting for inflation and quantity effects, Drezner *et al*. (1993:49) reported that the average cost growth on a sample of 197 completed defense programs was twenty percent, while the average cost overrun, the difference between the final budget and cost, was only eight percent. The twelve- percent difference may be attributable to modifications, and these modifications would almost certainly contribute to baseline volatility.

Cost overruns and baseline stability. Despite numerous assertions about a relationship between cost overruns and baseline volatility (e.g., Archibald 1976:196, Augustine 1983:110-140 and 1990:2, Gilbraeth 1986:37, Harroun *et al.* 1993:10), we could find no research which directly tested for a relationship. Furthermore, we could find only a few studies that indirectly tested for the effects of baseline changes. Each of these evaluated the effect of baseline changes on some parameter related to cost overruns, including the cost performance index, the estimate at completion, and recoveries from cost overruns. As indicated in Table 1, the results of these studies were mixed.

Based on their analyses of hundreds of completed defense contracts, Christensen and Payne (1991) and Christensen and Heise (1992) reported empirical evidence in support of a popular rule-of-thumb, known as the "ten percent rule:"

> *After a contract is twenty percent complete, the cumulative cost performance index (CPI) does not change by more than ten percent, and in most cases it only worsens.*

This important rule-of-thumb is a benchmark used in evaluating the reasonableness of the estimated final cost of a defense contract, known as the "estimate at completion" (EAC). Both studies reported that the results were not sensitive to numerous factors, including baseline changes. However, neither study specifically focused on the potential relationship between baseline changes and cost overruns.

Based on an analysis of 321 defense contracts, Terry and Vanderburgh (1993) reported that the EAC computed using the cumulative schedule-cost index (SCI), the product of cumulative cost and schedule performance indices, was the most accurate index-based EAC formula. They also reported that the relative accuracy of numerous index-based formulas which they examined was sensitive to "major baseline changes, " defined as an allocation of budget in excess of the contract's budget baseline. Although this condition, known as "over-target baseline," is discouraged by defense policy, it is not uncommon. To the extent that the accuracy of the EAC may influence the final cost of a contract, these results suggest that major changes to the budgetary baseline may influence a cost overrun.

In 1994 Elkinton and Gondeck tried to improve the accuracy of index-based EAC formulas by including an adjustment factor for cost growth in the budgetary baseline. Unfortunately, the results showed that these adjusted EAC formulas were less accurate than the unadjusted formulas, thus suggesting that changes to the baseline may not affect cost overruns.

Also in 1994, Pletcher and Young examined management actions that may have contributed to the recoveries for early cost overruns reported on defense contracts. As may be expected, recoveries are rare. In their sample of 303 contracts, only eight successfully recovered from early cost overruns. The average change in the performance measurement baseline for the eight contracts was 25%, compared to 67% for the rest of contracts. Thus, Pletcher and Young (1994:43) concluded that:

> *The single concrete measure that indicated a possible cause for recovery on the successful contracts was the stability of the performance measurement baseline (PMB). The successful contracts experienced a significantly more stable PMB than the remaining contracts. A stable PMB permits more accurate planning and leads to more effective use of resources.*

An important step in controlling defense cost overruns is to understand their causes. Given the rather inconsistent results described above, testing for a relationship between cost overruns and baseline instability on defense contracts seems justified.

METHODOLOGY

To determine whether cost overruns are related to baseline stability, we tested cost data from completed defense contracts with two statistical models. Each model used different measures for cost overruns and baseline stability. We used two models, each with different measures, because we wanted to test whether the results depended on the statistical models. A brief description of each model and the database follows.

<u>Two-factor ANOVA</u>. The first model tested for a relationship between cost overruns and baseline stability using a two-factor analysis of variance (ANOVA) model of the form:

$$SCI = \mu + \alpha + \beta + \alpha\beta + \epsilon \qquad (1)$$

where the Schedule Cost Index (SCI) is a measure of cost overrun, μ is the population average, α is the variation in SCI due to the stage of completion on the contract, β is the variation due to the number of significant changes to the contract baseline, is variation due to interaction between completion stage and baseline stability, and ϵ is the variation in SCI due to all other possible factors.

The SCI was chosen as the measure of cost overrun because research has shown it to be one of the best predictors of the final cost of a defense contract (Christensen *et al.* 1995, Terry and Vanderburgh 1993). As the product of cumulative CPI and cumulative Schedule Performance Index (SPI), the SCI captures both cost and schedule status, and is particularly useful in the early-to-middle stages of a contract when accurate estimates are most needed.

The cumulative Cost Performance Index (CPI) is defined as the cumulative Budgeted Cost of Work Performed (BCWP) divided by the cumulative Actual Cost of Work Performed (ACWP). When the index is less than one, a cost overrun is indicated. The Schedule Performance Index (SPI) is defined as BCWP divided by the Budgeted Cost of Work Scheduled (BCWS). When the SPI is less than one, a behind schedule condition is indicated. As a contract approaches completion, the SPI goes to unity and the SCI approaches the CPI.

The number of *significant* changes to the Contract Budget Base (CBB) was chosen as the measure of baseline stability. The CBB includes the budget for the known work on the contract, termed the Budget at Completion (BAC), and the budget for Management Reserve (MR). A significant change was defined as any individual or cumulative change in the CBB greater than ten percent. For example, suppose that a twelve month contract is awarded for $100 million, excluding profit or fee. In the second month, the contract is modified to increase several minor performance specifications, and the CBB is raised to $120 million. This is one change. Now suppose that in the following month a contractor credit proposal is accepted to waive several military standards, resulting in cost savings of $6M, or 5 percent. By itself, this change is too small to be considered a significant change. However, six months later the contractor receives $6M to conduct a subsystem design review, which had been deleted in the previous modification. Although the last two changes result in no effective change to the baseline, together they are significant because they add to over ten percent. Thus, in this example, the contract is changed significantly twice.

We tested for a *statistically* significant relationship between SCI and number changes to the baseline by setting the level of significance to be 5 percent.[1] The level of significance pertains to the chance of committing a Type I error: concluding that there is a relationship between the SCI and the number of changes to the baseline when there is no relationship. In practice, a value of 5 percent is used for the level of significance (Shao 1976).

The primary factor in the model (number of changes) was evaluated at five levels (no changes, two changes, three changes, and four or more changes).[2] The secondary factor in the model (percent complete) was included because prior research (e.g., Christensen 1993) shows that the magnitude of a cost overrun can depend on the stage of a project's completion. We evaluated this factor at four levels of percent complete (10-25, 25-50, 50-75, and 75-100). Contract data before the 10 percent completion point were not considered reliable.

Linear Regression Analysis. The second model tested for a relationship between cost performance and baseline stability using linear regression of the form:

$$CPI_L = \beta_0 + \beta_1 (CV_{PMB}) + \varepsilon \qquad (2)$$

where CPI_L is the last reported cumulative CPI, β_0 and β_1 are the regression coefficients, CV_{PMB} is the Coefficient of Variation of the Performance Measurement Baseline, and ε is the error term. The last reported cumulative CPI was chosen as a measure for cost overrun because it is a good predictor of final cost performance for contracts beyond the 75 percent completion point (Christensen, *et al.* 1995).[3]

Viewing the budget baseline as a random variable, several measures of volatility were considered, including variance, standard deviation, and coefficient of variation. Of these, the coefficient of variation of the baseline was most suitable because it measures *relative* dispersion. The variance and standard deviation are absolute measures of dispersion, and would be biased by the different sizes of the contracts compared in this study. The equation for the coefficient of variation is

$$CV_{PMB} = \sigma_{PMB} / \mu_{PMB} \qquad (3)$$

where σ_{PMB} and μ_{PMB} are the standard deviation and mean of the Performance Measurement Baseline (PMB). The PMB is defined as the CBB minus management reserve budget.

In this model, the regression coefficient (β_1) was used to measure statistical significance. If β_1 was significantly different from zero, we concluded that there was a relationship between baseline variability and cost overruns. As with the first model, the level of significance was set at 5 percent.[4]

Finally, the sensitivity of the results to military service (army, navy, air force), contract type (fixed price or cost reimbursable) and buying activities (e.g., air, ground, sea, and electrical systems) were evaluated by running regressions on the subgroups.

The Database. We used the Defense Acquisition Executive Summary (DAES) database for the analysis. It contains detailed cost and schedule performance data on hundreds of major defense acquisition programs.[5] Each program has dozens of contracts involving the development or production of military weapon systems (e.g., aircraft, ships, missiles). An example of a defense acquisition program in the database is the Air Force program to develop and produce the C-17. The C-17 program has many R&D and production contracts. Some of the contracts are complete; others are on-going.

Since the 1960s defense contractors have prepared monthly Cost Performance Reports (CPR) and submitted them to government program offices for review and analysis. Each program office summarizes the CPR data into a DAES report which is submitted at least quarterly to the Office of the Under Secretary of Defense where it is maintained in electronic form. The 1995 version of the database used here included 378 programs with cost performance data from over 1,000 defense acquisition contracts. In addition to cost and schedule data, the database has fields that describe each contract by type (cost reimbursable, fixed-price), phase (R&D, production), military service (army, navy, air force), and buying activity (e.g., air, ground, sea, electronic).

To the extent possible, all of the contracts in the DAES database were included. For both models, contracts less than ten percent complete were eliminated because it often takes at least that long for a contract's budgetary baseline to be established, and for a contractor to be found compliant the DOD's management standards, known as "Earned Value Management Systems Criteria."[6] Until each of these conditions is met, the reliability of the performance data is dubious. This elimination left 669 contracts from 165 programs available for the ANOVA model. Because the focus of the regression model was on mature contracts with complete histories, eliminating contracts that were less than 75 percent complete or had not reported performance prior to being 25 percent complete further reduced the sample. This left a data set of contracts of 401 contracts covering 131 programs.

Limitations. Using the DAES database had limitations that affected the validity our results. Although criteria-compliant contractors prepared the data, the degree of compliance and oversight is hardly constant across all contractors and the services that manage them. Also, only major defense acquisition programs were included. Smaller programs that did not meet the dollar thresholds for CPR-reporting and most firm-fixed-price contracts were excluded.[7] Finally, the operational measures used for cost overrun (SCI and CPI) and baseline stability (number of significant changes and the coefficient of variation) may not adequately capture the meaning of these terms as used by those who assert that a relationship exits. Given these limitations, generalizing the results to on-going and future contracts may be inappropriate. However, we have attempted to test the results for sensitivity to these limitations by (1) testing for differences across services, and buying activities, and (2) by using two statistical models. In addition, the assumptions of each statistical model were rigorously tested. A description of the results follows.

RESULTS

Overall, we found no statistical relationship between baseline volatility and cost overruns. This was true for each model. These results were not sensitive to the managing service (army, navy, air force), type of contract (fixed price, cost-reimbursable), nor the buying activity (e.g., air, ground, sea, electronics).

Table 2 summarizes the results of the ANOVA model on 669 contracts. As indicated by the small F statistics, the Schedule Cost Index, the measure of cost performance in the model, was not affected by the stage of completion (factor A), the number of significant changes (factor B), nor the interaction of factors A and B.[8]

Table 3 summarizes the results of the linear regression model on 401 contracts. In the model, the last Cost Performance Index (CPI_L) was the dependent variable, and the coefficient of variation of the Performance Measurement Baseline (PMB) was the independent variable. As indicated by the small t statistics, the regression coefficient ($\square_1$) for each category examined was not significantly different from zero.[9]

The two exceptions were an Air Force product center (F04704) with 64 contracts, and an Army product center (DAAB07) with 7 contracts.[10] For each of these, the slope of the regression equation coefficient was significantly positive, indicating that more variability in the contract baselines was associated with favorable cost performance—a result counter to the hypothesized relationship. For the Air Force product center, 51 of the 64 contracts were fixed price, and retained a significantly positive regression coefficient. The remaining 13 contracts were cost reimbursable, and showed no relationship.

IMPLICATIONS

The results indicate that contracts do as well under conditions of change as they do under conditions of stability. The observation that volatile contracts do no worse than stable ones could be attributed to an appropriate amount of management attention directed to contracts with significant change. This conclusion would be welcome in the defense industry after decades of allegations of mismanagement (e.g., Garland 1990, Ghosh 1995, Fox 1988). However, to be complete, alternative conclusions should be explored.

There are at least two possible explanations of why a volatile contract would not overrun more than a stable one. The first is the possibility of funding an overrun; the other is the effect of "buying in."

Funding an overrun. When a change in the work on a contract occurs, an equitable adjustment should be made. If the different terms and conditions result in increased cost to the contractor, then the contractor is due additional budget to accept those terms and conditions. If new terms decrease the scope (e.g., quantity) or change the period of performance, then the government may be due a reduction in contract price. The converse is also true. When a contractor receives more or less money for work, then the nature of that work should have also changed. If, however, a contractor who is experiencing a cost overrun excessively overvalues the cost of new work, the contractor could be attempting to capture excess funds to cover losses.

Buying-in. A related situation can occur when a contractor is under market pressure to reduce the estimate of the cost of a new project. This pressure could be either from competition or from the suspicion that if the full cost of a project were known, then it would not be funded. The practice of intentionally underestimating the magnitude of a contract is termed "buying in." A common expression heard in defense circles is "bid 'em low and watch 'em grow," referring to the opportunity to regain losses on an under-budgeted project from overpriced changes (Augustine 1983). If true, this characteristic would also explain the results identified in this study. An unstable contract would have more opportunities to "watch 'em grow."

Cost growth versus cost overrun. It is also important to recognize a difference between cost growth and cost performance. While some studies focus on cost growth (e.g., Drezner *et al.* 1993, Tyson *et al.* 1994, Hough 1992, Jarvaise *et al.* 1996), others focus on cost overruns (e.g., Christensen *et al.* 1995, Christensen 1994). For example, Drezner *et al.* (1993) do not separate cost performance (overruns or under-runs) from cost growth. By saying that a program grew by 20 to 30 percent, Drezner *et al.* (1993) are not saying that a contractor overspent the authorized budget by that much. Rather, they are saying that the final bill for the program increased. Thus, when the literature asserts that cost growth is sensitive to baseline changes, it may simply mean that changes usually increase the authorized budget in the baseline, but do not necessarily lead to cost overruns.

Other factors. A more practical and less sinister alternative also exists. Early in the contracting process, proposals are often made with a great deal of uncertainty. Work may only be vaguely analogous to previous experience. New technology may be envisioned which could reduce the cost, or if troublesome, increase it. Personnel projections are difficult to make when the lure of the next big project may draw away the assembled team of experts. Corporate dynamics can also be hardly anticipated. The project may be in a division swallowed up into an entirely different contractor, or divested to make it on its own. Also important is the future business base. If a contractor fails to win enough future work, the cost of supporting their infrastructure would be born by fewer and fewer projects. These higher overhead rates could then scare away potential new work, perpetuating the problem, and driving up the cost of existing work. All of these factors bear on the total cost of a project, for which a contract price is based on the most tenuous of estimates. Combined with the pressures mentioned previously, a winning bidder would almost always estimate on the low end of a wide distribution of likely costs.

Having agreed to a price that ensured a 50% or greater chance of an overrun at completion, the contractor would welcome additional incremental changes. These incremental changes, often well after the work has been dissected and detailed, are much easier to estimate. The ultimate example would be to add one more unit to an order after a few units had already been produced, or vice versa. The cost of the new or last unit would be predictable with a higher degree of certainty. Therefore, the chance of under-running on the new work, or taking advantage of well understood costs to pad the estimate upward would be very high. Making this a virtual certainty is that most modifications are executed without competition. Competition is an important reason the contractor did not pad their original estimate to reduce risk in the first place.

CONCLUSION

It is evident from the range of possible explanations that further research in this area is needed. Specifically, management practices relative to change management, if better understood, could lead to a sharing of lessons learned. The programs that experienced high volatility without suffering overruns may have developed techniques to cope with their environment. Similarly, programs enjoying stable baselines and negative cost performance may have discovered pitfalls that should be avoided. In particular, analysis of the 51 fixed price contracts from activity F04704 may reveal peculiarities or techniques used to ensure the best cost performance in the study population, averaging 1% under budget in their final quarter and a positive relationship with contract changes. Finally, if unethical practices exist, they should be discovered and eradicated. Whatever the actual source of these results, the program management community stands to gain from a better awareness and understanding of the variety of situational realities versus dependence on anecdotal evidence and popular opinion.

REFERENCES

Archibald, Russell D. (1976). *Managing High Technology Programs and Projects.* New York: John Wiley & Sons.

Augustine, Norman R., *et al.* (1990). *Report of the Advisory Committee on the Future of the U.S. Space Program.* U.S. Government Printing Office, Washington DC.

Christensen, David S. (1994). "Cost Overrun Optimism: Fact or Fiction." *Acquisition Review Quarterly 1*:23-38 (Winter).

Christensen, David S. (1996). "Project Advocacy and the Estimate at Completion Problem." *Journal of Cost Analysis*, pp. 35-60 (Spring).

Christensen, David S. and Kirk Payne (1992). "Cost Performance Index Stability- Fact or Fiction?" *Journal of Parametrics* 10:27-40 (April).

Christensen, David S. and Scott R. Heise (1993). "Cost Performance Index Stability." *National Contract Management Journal, 25(1)*: 7-25.

Christensen, David S. (1993). "An Analysis of Cost Overruns on Defense Acquisition Contracts." *Project Management Journal 24*:43-48 (September).

Christensen, David S., Richard Antolini, and John McKinney (1995). "A Review of Estimate At Completion Research." *Journal of Cost Analysis*, pp. 41-62 (Spring).

Czelusniak, Daniel P. and Philip D. Rodgers (1997). "Program Stability - The Kaminski Initiative." *Program Manager 26*:59-64 (September-October).

Department of Defense. DOD Regulation 5000.2-R (1996). "Mandatory Procedures for Major Defense Acquisition Programs (MDAPs) and Major Automated Information Systems (MAIS) Acquisition Programs."

Drezner, Jeffrey A., J. M. Jarvaise, R. W. Hess, P. G. Hough, and D. Norton (1993). *An Analysis of Weapon System Cost Growth.* MR-291-AF. Santa Monica, CA: RAND Corporation.

Elkinton, Timothy L. and Todd J. Gondeck (1994). *A Study to Determine if Newly Developed Budget at Completion Adjustment Factors Improve the Success of Prevalent Index-Based Estimate at Completion Techniques*, Masters thesis, Wright Patterson AFB, OH: Air Force Institute of Technology (September).

Fox, J. Ronald (1988). *The Defense Management Challenge: Weapons Acquisition*, Boston, MA: Harvard Business Press.

Garland, Howard (1990). "Throwing Good Money After Bad: The Effect of Sunk Cost on the Decision to Escalate Commitment to an Ongoing Project." *Journal of Applied Psychology 75*: 728-731 (December).

General Accounting Office (1997). "Major Acquisitions - Significant Changes Underway in DOD's Earned Value Management Process." GAO/NSIAD-97-108 (May).

Ghosh, Dipankar (1995). "Throwing Good Money After Bad." *Management Accounting 77*: 51-54 (July).

Gilbraeth, Robert D. (1986). *Winning at Project Management: What Works, What Fails and Why.* New York: John Wiley & Sons.

Harroun, William P., Jeffrey S. McCarthy, and Dan D. Epperson (1993). "The Importance of Scope Management." *Proceedings of the 5th Annual National Cost Schedule Performance Management Conference*, Washington DC (24 October).

Hough, Paul G. (1992). *Pitfalls in Calculating Cost Growth From Selected Acquisition Reports*, N-3136-AF, Santa Monica, CA: RAND Corporation.

Jarvaise, J. M., J. A. Drezner, D. Norton (1996). *The Defense System Cost Performance Database*, MR-625-OSD, Santa Monica, CA: RAND Corporation.

Mayer, Kenneth R. (1991). *The Political Economy of Defense Contracting*, Yale, New Haven CT.

Peck, Merton J. and Frederick M. Scherer (1962). *The Weapon Acquisition Process: An Economic Analysis*, Boston, MA: Harvard University Press.

Pletcher, John M. and Jane C. Young (1994). *An Evaluation of Department of Defense Contractors' Cost Performance*, Masters thesis. Wright-Patterson AFB, OH: Air Force Institute of Technology (September).

Shao, Stephen P. (1976). *Statistics for Business and Economics*. Columbus, Ohio: Charles E. Merrill Publishing Company.

Terry, Mark F. and Mary M. Vanderburgh (1993). *An Analysis of Estimate at Completion Models Utilizing the Defense Acquisition Executive Summary Database*, Masters thesis, Wright-Patterson AFB, OH: Air Force Institute of Technology (September).

Tyson, Karen W., Bruce R. Harmon, and Daniel M. Utech (1994). “Understanding Cost and Schedule Growth in Acquisition Programs.” Alexandria, Virginia: Institute for Defense Analysis (July).

Winer, B. J. (1971). *Statistical Principles In Experimental Design*. Second Edition. New York: McGraw-Hill.

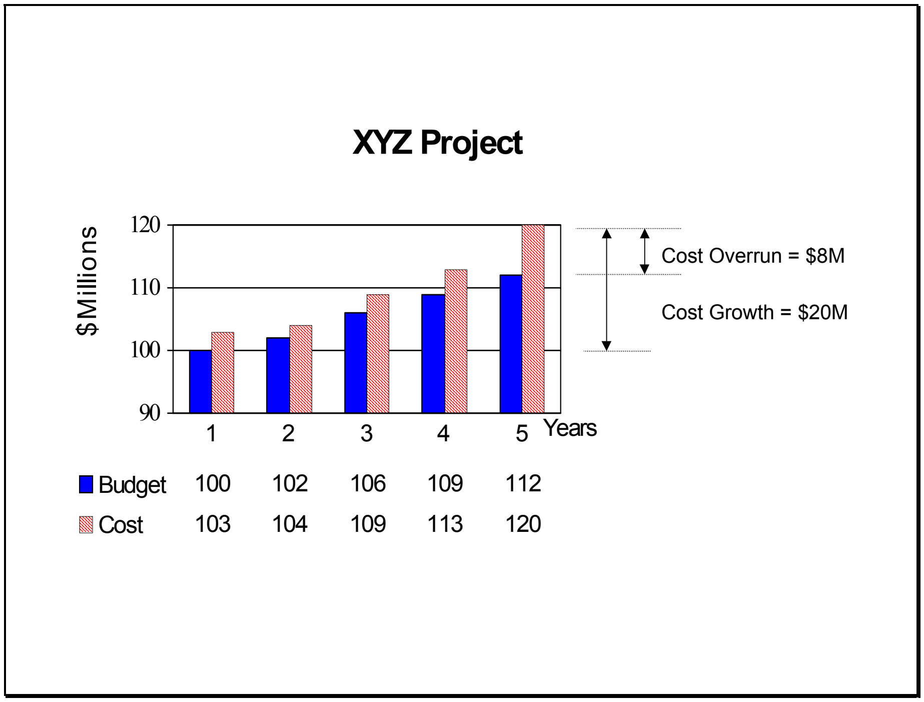

Figure 1. Cost Growth versus Cost Overrun.

TABLE 1

COST OVERRUN STUDIES EXAMINING BASELINE EFFECTS

Author (year)	***Topic of study***	***Sensitive to baseline changes?***
Christensen and Payne (1991)	Cost performance index stability	Not sensitive
Christensen and Heise (1992)	Cost performance index stability	Not sensitive
Terry and Vanderburgh (1993)	Estimate at completion accuracy	Sensitive
Elkinton and Gondeck (1994)	Estimate at completion accuracy	Not sensitive
Pletcher and Young (1994)	Cost overrun recoveries	Sensitive

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE

Sources of Variation	*Sums of Squares*	*Degrees of Freedom*	*Mean Square*	*F*	*F critical (α=.05)*
Factor A: Quarters	0.1099	3	0.0366	1.251	2.618
Factor B: Changes	0.0890	4	0.0223	0.760	2.385
Interaction	0.5914	12	0.0493	1.683	1.767
Error	19.561	668	0.0293		

TABLE 3

SUMMARY OF REGRESSION ANALYSIS

Sample		***CPI_L***		***CV_{PMB}***		***Slope***		
Category	***n***	***Mean***	***Std Dev***	***Mean***	***Std Dev***	***β_1***	***Std Error***	***t***
All AF	151	0.96	0.11	0.36	0.58	0.004	0.015	0.271
All Navy	142	0.94	0.11	0.32	0.42	-0.009	0.219	-0.407
All Army	108	0.92	0.12	0.62	0.72	-0.006	0.016	-0.364
Air Force Sources								
F04704	64	1.00	0.10	0.32	0.53	0.059	0.023	2.62*
F19628	15	1.00	0.09	0.32	0.42	-0.011	0.060	-0.178
F33657	44	0.93	0.10	0.29	0.48	-0.041	0.033	-1.246
F04701	18	0.89	0.09	0.36	0.43	0.005	0.055	0.092
Navy Sources								
N00030	15	0.97	0.04	0.43	0.49	-0.015	0.019	-0.769
N00024	88	0.96	0.11	0.32	0.46	-0.006	0.026	-0.229
N00019	34	0.90	0.12	0.29	0.28	-0.070	0.073	-0.964
Army Sources								
DAAH01	42	0.95	0.09	0.49	0.60	-0.001	0.024	-0.409
DAAK30	9	0.93	0.20	0.67	0.56	0.041	0.133	0.311
DAAK50	14	0.93	0.09	0.66	0.71	-0.001	0.037	-0.033
DAAB07	7	0.89	0.11	0.91	0.51	0.169	0.064	2.630*
DAAK40	9	0.89	0.09	0.58	0.69	-0.084	0.038	-2.219**
DAAJ01	13	0.89	0.12	1.10	1.25	-0.005	0.028	-0.171
Fixed Price								
F04704	51	1.01	0.10	0.35	0.100	0.057	0.024	2.379*
N00024	73	0.96	0.12	0.24	0.12	-0.041	0.037	-1.113
DAAH01	20	0.94	0.10	0.57	0.70	0.001	0.035	0.036
F33657	39	0.94	0.10	0.30	0.50	-0.041	0.033	-1.245
N00019	18	0.93	0.11	0.26	0.24	0.046	0.116	0.394
DAAK50	12	0.91	0.08	0.70	0.76	0.010	0.034	0.302
F04701	15	0.88	0.09	0.35	0.46	0.021	0.055	0.379
Cost Plus								
N00030	14	0.97	0.04	0.36	0.43	-0.027	0.024	-1.163
F04704	13	0.97	0.07	0.19	0.08	-0.128	0.259	-0.496
DAAH01	22	0.96	0.08	0.41	0.50	-0.025	0.037	-0.644
N00024	14	0.95	0.08	0.73	0.66	0.059	0.031	1.910**
N00019	15	0.86	0.12	0.32	0.33	-0.127	0.091	-1.390

* $p < 0.05$ ** $p < 0.10$

[1] In ANOVA, a factor is significant when its F statistic exceeds a critical value. The critical value is determined by the desired level of significance and by the sample size. If the F statistic for a factor is more than the critical value, the null hypothesis (e.g., the SCI is the same for any number of changes to the baseline) is rejected at the pre-specified level of significance. When the null hypothesis is rejected, the alternative hypothesis (e.g., the SCI is not the same for any number of baseline changes) is accepted.

[2] Because the number of observations was not equal in each group (varying from 3 to 134 observations per group), a special case of ANOVA which adjusts for unequal sample sizes was used (Winer 1971:210). The assumptions regarding the model's error term were tested using the chi-square test of normality and Bartlet's test for equal variance at the 5% level of significance (Winer 1971).

[3] As with the ANOVA model, the final cost performance of many contracts had to be predicted because performance reporting is often discontinued after the 90 percent completion point. If reports were submitted until a contract was 100 percent complete, the CPI_L is the actual final cost performance on the contract, because cumulative BCWP will equal the final budget, termed the Budget at Completion (BAC), by definition. To reduce possible bias created by incomplete or partially reported contracts, only contracts reporting data from before 25 through greater than 75 percent complete were included.

[4] In regression analysis, a t-statistic is typically used to determine if a coefficient is significantly different from zero. The t-statistic is calculated by dividing the standard error of the coefficient into the coefficient. The observed t-statistic is then compared to a critical value. The critical value depends on the desired level of significance and the sample size. When the observed t-statistic is larger than the critical value, the coefficient is significantly different from zero, and the null hypothesis (e.g., the final CPI is the same for any amount of baseline variability) is rejected at the pre-specified level of significance. When the null hypothesis is rejected, the alternative hypothesis (e.g., the final CPI is not the same for any amount of baseline variability) is accepted.

[5] Most of the programs in the database are acquisition category 1, defined as "a major defense acquisition program with estimated expenditures of over $355 million in research, development, test and evaluation, or over $2.135 billion in procurement (in fiscal year 1996 dollars.)." (DOD Regulation 5000.2R, Section C7, page 3).

[6] These criteria, formerly termed "Cost/Schedule Control Systems Criteria," were initially intended to be flexible management guidelines for contractors. Through the years, however, the criteria were over-implemented with audit-like compliance reviews. Eventually, the criteria became an administrative burden for the contractor and the government. In 1996 the criteria were revised by industry, adopted by the government, and re-named "Earned Value Management Systems Criteria." (GAO 1997).

[7] The dollar thresholds for CPR reporting and criteria compliance have gradually increased since 1967 when the criteria where initially implemented. Presently, the criteria are required on R&D contracts with costs over $70 million and on procurement contracts with costs over $300 million (1996 fiscal year constant dollars).

[8] At the 5% level of significance, the F statistics for each factor were less than the critical values. Thus, the null hypothesis that the SCI was equal across number of changes to the baseline could not be rejected. Likewise, the null hypothesis that the SCI was equal across quarters could not be rejected.

[9] At the 5% level of significance, the t-statistics for most of the subgroups were less than the critical values. Thus, the null hypothesis that the final CPI was the same for any amount of baseline variation could not be rejected.

[10] A single asterisk in Table 3 indicates that the probability of making a type I error (rejecting the null hypothesis when it is true) is less than 5% ($p<5\%$). Two asterisks indicate that the probability of making a type I error is less than 1% ($p<1\%$). Typically, when p is less than 1%, the observation is characterized as "highly significant" (Shao 1976).